

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re the Application of:

Mi-SUK LEE, ET AL.

Application No.:

Filed:

For: **METHOD OF ESTIMATING PITCH BY
USING RATIO OF MAXIMUM PEAK TO
CANDIDATE FOR MAXIMUM OF
AUTOCORRELATION FUNCTION AND
DEVICE USING THE METHOD**

Art Group:

Examiner:

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

REQUEST FOR PRIORITY

Sir:

Applicant respectfully requests a convention priority for the above-captioned application, namely:

COUNTRY	APPLICATION NUMBER	DATE OF FILING
Korea	2002-0061787	10 October 2002

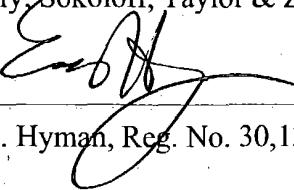
A certified copy of the document is being submitted herewith.

Dated: 7/27/03

12400 Wilshire Blvd., 7th Floor
Los Angeles, California 90025
Telephone: (310) 207-3800

Respectfully submitted,

Blakely, Sokoloff, Taylor & Zafman LLP


Eric S. Hyman, Reg. No. 30,139

**KOREAN INTELLECTUAL
PROPERTY OFFICE**

This is to certify that the following application annexed hereto is a true copy from the records of the Korean Intellectual Property Office.

Application Number:: Korean Patent Application 2002-0061787

Date of Application:: 10 October 2002

Applicant(s): Electronics and Telecommunications Research Institute

10 July 2003

COMMISSIONER

[Bibliography]

[Document Name] Patent Application
[Classification] Patent
[Receiver] Commissioner
[Reference No.] 0001
[Filing Date] 10 October 2002
[IPC] G10L

[Title] The pitch estimation algorithm by using the ratio of the maximum peak to candidates for the maximum of the autocorrelation function

[Applicant]
[Name] Electronics and Telecommunications Research Institute
[Applicant code] 3-1998-007763-8

[Attorney]
[Name] Young-pil Lee
[Attorney code] 9-1998-000334-6
[General Power of Attorney]
Registration No.] 2001-038378-6

[Attorney]
[Name] Hae-young Lee
[Attorney code] 9-1999-000227-4
[General Power of Attorney]
Registration No.] 2001-038396-8

[Inventor]
[Name] LEE, Mi Suk
[Resident]
Registration No.] 700310-2451057
[Zip Code] 305-350
[Address] 161 Gajeong-dong, Yusong-gu, Daejeon-city
Rep. of Korea
[Nationality] Republic of Korea

[Inventor]
[Name] HWANG, Dae Hwan
[Resident]
Registration No.] 630716-1010012
[Zip Code] 305-390
[Address] 103-905 Narae Apt., Jeonmin-dong, Yusong-gu, Daejeon-city
Rep. of Korea
[Nationality] Republic of Korea

[Request for
Examination]

Requested

[Purpose]

We file as above according to Art. 42 of the Patent Law, request
the examination as above according to Art. 60 of the Patent Law.

Attorney

Young-pil Lee

Attorney

Hae-young Lee

[Fee]

[Basic page]	20 Sheet(s)	29,000 won
[Additional page]	5 Sheet(S)	5,000 won
[Priority claiming fee]	0 Case(S)	0 won
[Examination fee]	8 Claim(s)	365,000 won
[Total]	399,000 won	

[Reason for Reduction] Government Invented Research Institution

[Fee after Reduction] 199,500 won

[Transfer of Technology] Allowable

[Licensing] Allowable

[Technology Training] Allowable

[Enclosures]

1. Abstract and Specification (and Drawings) 1 copy

대한민국 특허청

KOREAN INTELLECTUAL
PROPERTY OFFICE

별첨 사본은 아래 출원의 원본과 동일함을 증명함.

This is to certify that the following application annexed hereto
is a true copy from the records of the Korean Intellectual
Property Office.

출원번호 : 10-2002-0061787

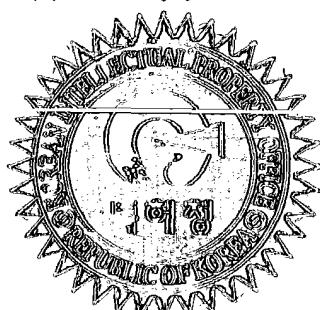
Application Number

출원년월일 : 2002년 10월 10일
Date of Application

OCT 10, 2002

출원인 :
Applicant(s)

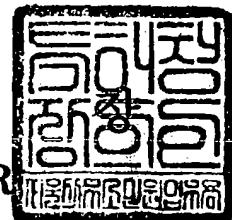
한국전자통신연구원
Electronics and Telecommunications Research Institu-



2003 년 07 월 10 일

특허청

COMMISSIONER



【서지사항】

【서류명】	특허출원서
【권리구분】	특허
【수신처】	특허청장
【참조번호】	0001
【제출일자】	2002. 10. 10
【국제특허분류】	G10L
【발명의 명칭】	상관함수의 최대값과 그의 후보값의 비를 이용한 피치 검출 방법 및 그 장치
【발명의 영문명칭】	The pitch estimation algorithm by using the ratio of the maximum peak to candidates for the maximum of the autocorrelation function
【출원인】	
【명칭】	한국전자통신연구원
【출원인코드】	3-1998-007763-8
【대리인】	
【성명】	이영필
【대리인코드】	9-1998-000334-6
【포괄위임등록번호】	2001-038378-6
【대리인】	
【성명】	이해영
【대리인코드】	9-1999-000227-4
【포괄위임등록번호】	2001-038396-8
【발명자】	
【성명의 국문표기】	이미숙
【성명의 영문표기】	LEE, Mi Suk
【주민등록번호】	700310-2451057
【우편번호】	305-350
【주소】	대전광역시 유성구 가정동 161번지
【국적】	KR
【발명자】	
【성명의 국문표기】	황대환
【성명의 영문표기】	HWANG, Dae Hwan
【주민등록번호】	630716-1010012

【우편번호】 305-390
【주소】 대전광역시 유성구 전민동 나래아파트 103동 905호
【국적】 KR
【심사청구】 청구
【취지】 특허법 제42조의 규정에 의한 출원, 특허법 제60조의 규정에 의한 출원심사를 청구합니다. 대리인
 이영필 (인) 대리인
 이해영 (인)
【수수료】
 【기본출원료】 20 면 29,000 원
 【가산출원료】 5 면 5,000 원
 【우선권주장료】 0 건 0 원
 【심사청구료】 8 항 365,000 원
【합계】 399,000 원
【감면사유】 정부출연연구기관
【감면후 수수료】 199,500 원
【기술이전】
 【기술양도】 희망
 【실시권 허여】 희망
 【기술지도】 희망
【첨부서류】 1. 요약서·명세서(도면)_1통

【요약서】

【요약】

본 발명은 대부분의 음성 코덱에서 사용하고 있는 오픈 루프(open-loop) 피치 검색 기의 성능을 향상시키기 위한 방법 및 그 장치에 관한 것이다. 본 발명에서 제안하고 있는 오픈 루프 피치 검출 장치는 perceptual weighting filtering 된 음성 신호로부터 정규화된 상관 함수를 구하는 상관함수 계산부, 상기 상관 함수 값들 중에서 최대값과 상기 최대값의 후보값들을 검출하는 최대 상관함수 및 그의 후보값 검출부, 상기 검출된 최대 상관함수와 그 후보 값들의 비를 계산하고, 소정의 문턱치보다 작은 값을 갖는 래그를 검출하는 피치 후보 결정부, 및 이전 프레임의 피치를 이용하여 상기 새로이 구한 피치 후보들과 최대 상관함수값을 갖는 래그중에서 피치를 선택하는 피치 검출부를 포함하여, 계산량이 적을 뿐만 아니라 기존의 알고리즘에 비해 배수(multiple) 혹은 역수(submultiple) 피치 오류가 적기 때문에 음성 코덱의 음질 향상에 기여할 수 있다.

【대표도】

도 5

【색인어】

상관함수, 피치, 음성코덱

【명세서】

【발명의 명칭】

상관함수의 최대값과 그의 후보값의 비를 이용한 피치 검출 방법 및 그 장치{The pitch estimation algorithm by using the ratio of the maximum peak to candidates for the maximum of the autocorrelation function}

【도면의 간단한 설명】

도 1은 CELP 음성 코덱의 부호화기에 대한 블록도이며,

도 2는 CELP 음성 코덱의 복호화기에 대한 블록도이며,

도 3은 여성 화자의 perceptual weighting filtering된 음성 신호와 정규화된 상관 함수 값을 설명하는 참고도이며,

도 4는 도 3의 d_{max} 와 d_x 의 상관 함수값이며,

도 5는 본 발명에서 제안하는 Open-loop 피치 검색기이며,

도 6은 최대 상관 함수 값을 갖는 래그를 피치로 검출했을 때 배수 피치 오류가 발생한 프레임에 대한 $K(d_x)$ 값의 분포도이며,

도 7은 남성 화자의 perceptual weighting filtering된 음성 신호와 정규화된 상관함수값을 설명하는 참고도이며,

도 8은 도 7에 있는 d_x 에 대한 $K(d_x)$ 값을 설명하는 참고도이다.

【발명의 상세한 설명】

【발명의 목적】

【발명이 속하는 기술분야 및 그 분야의 종래기술】

<9> 본 발명은 대부분의 음성 코덱에서 사용하고 있는 오픈 루프(open-loop) 피치 검색 기의 성능을 향상시키기 위한 방법 및 그 장치에 관한 것으로, perceptual weighting filtering된 음성신호의 상관함수로부터 구한 최대값과 그의 후보값의 비를 이용한 피치 검출 방법 및 그 장치에 관한 것이다.

<10> 현재 가장 널리 사용되고 있는 CELP(Code Excited Linear Prediction) 방식의 음성 코덱에서는 한 프레임의 음성 신호로부터 스펙트럼 포락선을 나타내는 LPC(Linear Prediction Coefficient)와 음성 신호의 주기적인 특성을 나타내는 피치 그리고 피치가 제거된 LPC 분석 필터의 잔여 신호를 모델링하기 위한 고정 코드북 파라미터 등을 추출하고, 다시 이들 정보를 이용하여 음성 신호를 합성한다.

<11> 도 1은 일반적인 CELP 음성 코덱의 부호화기에 대한 블록도이다. 도 1을 참조하면 전처리부(101)에서는 입력 음성신호를 band-pass filtering하고 pre-emphasis 하는 일반적인 전처리 과정이 수행된다. LPC 분석/양자화부(102)는 LP 계수를 구한 후 전송을 위해 이 계수들을 양자화 한다. 합성필터부(103)의 입력 신호는 고정 코드북(104)과 적응 코드북(105)으로 모델링 된다. 위 적응 코드북(105)은 이전 프레임의 여기 신호로 업데이트 되는 코드북이다. 피치 검색부 (106)에서는 perceptual weighting filter부(108)를 통과한 입력 신호와 가장 비슷한 신호를 갖는 래그를 적응 코드북에서 검색하고 이 피치 검색부(106)에서 찾아진 래그를 피치라고 한다. 이때 적응코드북 검색에 많은 계

산량이 필요하기 때문에 먼저, open-loop 검색을 통해 대략적인 피치값을 구하고 이 값 주변의 한정된 래그에 대해서만 적응코드북을 검색한다. 고정 코드북 검색(107)기에서는 피치 정보가 제거된 LPC 분석 필터의 잔여 신호를 가장 잘 모델링 하는 고정 코드북 인덱스를 구한다. 이렇게 고정 코드북 인덱스와 피치 래그가 검출되면 각각의 코드북에 대한 이득을 계산하고 전송을 위해 양자화 한다(109).

<12> 도 2는 CELP 음성 코덱의 복호화기에 대한 블록도이다. 위 CELP 음성 코덱의 복호화기에서는 부호화기에서 검출된 파라미터를 이용하여 음성 신호를 합성한다. 부호화기에서 사용한 것과 같은 고정 코드북(201)과 적응 코드북(202)을 이용하여 재생된 여기 신호가 합성 필터(203)를 통과하게 되면 음성 신호가 합성되는데, 인간의 청각적 특성을 반영한 후처리 필터(204)를 사용하여 합성된 음성의 품질을 향상시킨다.

<13> 일반적으로 피치 검색기(106)는 계산량을 줄이기 위해 open-loop 피치 검색기와 closed-loop 피치 검색기로 구성된다. Open-loop 피치 검색기에서는 perceptual weighting filter를 통과한 음성 신호를 기반으로 하여 최대 상관도를 갖는 래그를 피치로 선택한다. 이때 실제 피치의 배수 혹은 역 수되는 래그를 피치로 검출하는 오류가 발생할 수 있다. 특히, 실제 피치의 배수 form를 피치로 검출하는 오류가 많이 발생한다. Closed-loop 검색기에서는 open-loop 검색기에서 구한 피치 주변의 한정된 래그에 대해서만 분석에 의한 합성 방식으로 좀 더 정확한 피치를 찾는다. 따라서 open-loop 피치 검색기에서 실제 피치의 배수나 역수를 피치로 검출한다면 이는 closed-loop 피치 검색기에서 복원할 수 없는 오류로, 음성 코덱의 합성음의 품질을 저하시키는 중요한 원인이 된다. 그러므로 open-loop 피치 검색기에서는 계산량이 적은 간단한 방법으로 피치 검출을 하되, 실제 피치의 배수나 역수를 피치로 검출하는 오류가 없어야 한다.

<14> Open-loop 피치 검색기의 오류를 줄이기 위해 여러 가지 알고리즘들이 제안되어 사용되고 있는데, 기존 음성 코덱에서 사용하고 있는 open-loop 피치 검색 방법은 크게 아래와 같이 두 가지로 나누어 볼 수 있다.

<15> ITU-T G.729 및 GSM EFR에서는 사용하고 있는 open-loop 피치 검색기에서는 피치 검색을 위한 래그의 범위를 삼 등분하여 각각에서 최대의 상관도를 갖는 래그를 찾은 후, 배수 피치 오류의 발생을 막기 위해 낮은 쪽 범위의 상관함수에 가중치를 주어 최대값을 갖는 래그를 피치로 검출한다. 이 방법은 화자에 따라 성능의 차이를 보일 수 있는데, 특히 여성화자에 비해 피치주기가 긴 남성화자의 경우 실제 피치의 역 수를 피치로 검출하는 오류가 발생할 수 있다.

<16> 3GPP 및 ITU-T에서 새로운 광대역 음성 코덱의 표준으로 선정한 AMR-WB에서는 이전 프레임의 피치를 이용한 피치 검출 알고리즘을 사용하고 있다. 여기서 사용하고 있는 피치 검색기에서는 낮은 래그의 상관함수에 가중치를 주고, 만일 현재 프레임이 유성음이라면 이전 프레임의 피치 값 근처 래그에 있는 상관함수에도 가중치를 준다. 이때 이전 프레임의 피치 값은 이전 다섯 프레임의 피치 값으로 median filtering 하여 사용한다. 이 피치 검출 방법의 성능은 이전 프레임의 피치값의 정확도에 의해 좌우되는데, 만일 이전 프레임의 피치가 현재 프레임의 피치의 배수가 될 경우에는 배수 피치 오류가 발생할 수 있다. 예를 들어, 무성음이나 천이 구간 근처의 유성음 구간에서 이전 프레임의 피치가 현재 프레임의 실제 피치의 배수였다면, 상관함수가 피치의 배수마다 피크를 갖기 때문에 실제 피치의 배수 래그에 해당하는 상관함수값에 가중치를 주게 되므로 이 배수 래그를 피치로 검출하는 배수 피치 오류가 발생 할 수 있는 문제점이 있다.

【발명이 이루고자 하는 기술적 과제】

<17> 본 발명이 이루고자 하는 기술적 과제는 음성신호의 최대 상관 함수값과 그의 후보값의 비를 이용하여 종래의 방법보다 정확한 피치를 검출하는 방법 및 그 장치를 제공하는데 있다.

【발명의 구성 및 작용】

<18> 상기 문제점을 해결하기 위해 본 발명에서 제안하고 있는 오픈 루프 피치 검출장치는 퍼셉추얼 웨이팅 필터링(PERCEPTUAL WEIGHTING FILTERING)된 음성신호로부터 정규화된 상관함수를 구하는 상관함수 계산부, 상기 상관함수 값을 중에서 최대값과 상기 최대값의 후보값들을 검출하는 최대 상관함수 및 최대 상관함수 후보값 검출부, 상기 검출된 최대 상관함수값과 그 후보 값들의 비를 계산하고 소정의 문턱치보다 작은 값을 갖는 래그를 검출하는 피치 후보 결정부, 및 상기 음성신호의 이전 프레임의 피치 값을 이용하여 상기 검출된 피치 후보와 최대 상관함수값을 갖는 래그 중에서 피치를 검출하는 피치 검출부를 포함하는 것을 특징으로 한다.

<19> 상기 오픈 루프 피치 검출 장치의 상기 상관함수 계산부는 다음 수학식

$$R(d) = \frac{\sum_{n=0}^{N-1} s_w(n-d)s_w(n)}{\sqrt{\sum_{n=2}^{N-1} s_w(n-d)^2}}$$

을 통하여 구하여지고,

<21> 여기서 $R(d)$ 는 상기 정규화된 상관함수, $s_w(n)$ 는 상기 perceptual weighting filtering된 음성신호, d 는 래그, d_L 과 d_H 는 최소와 최대 래그, N 은 피치 검색을 위한 윈도우 크기를 의미한다.

<22> 또한 상기 오픈 루프 피치 검출 장치의 상기 최대 상관함수값 및 그의 후보값 검출부는 상기 정규화된 상관함수값 중에서 가장 큰 값을 상기 최대값으로하고, 그 최대 값이 찾아지기 전까지의 과정에서 발생되었던 국부적인 최대값을 상기 최대값의 후보들로 선정하는 것을 특징으로 할 수 있다.

<23> 또한 상기 오픈 루프 피치 검출 장치의 상기 피치 후보 결정부는 상기 최대값의 후보들 각각에 대해 $K(d_x)$ 값을 다음 수학식

<24> $K(d_x) = \text{alpha} K_{\text{lag}}(d_x) + (1 - \text{alpha}) K_{\text{corr}}(d_x)$, $x=1,2,\dots,1$ 을 통하여 구하고, $K(d_x)$ 들 중 소정의 문턱치보다 작은 값을 구하되,

<25> 여기서 alpha 는 소정의 가중치 값,

<26> $K_{\text{lag}}(d_x)$ 는 다음 수학식

<27> $K_{\text{lag}}(d_x) | [d_{\text{max}}/d_x + 0.5] - d_{\text{max}}/d_x |$ 을 통하여 구하여 지는 값,

<28> d_x 는 상기 후보들 중 어느 하나의 래그 값,

<29> $K_{\text{corr}}(d_x)$ 값은 다음 수학식

<30> $K_{\text{corr}}(d_x) = |1 - R(d_{\text{max}})/R(d_x)|$ 을 통하여 구하여 지는 값

<31> i 은 d_{max} 이전에 나타났던 최대값 후보들의 수인 것을 특징으로 할 수 있다.

<32> 또한 상기 오픈 루프 피치 검출 장치의 상기 피치검출부는 이전 프레임에서 구한 피치를 이용하여 상기 피치 후보값과 최대 상관함수값을 갖는 래그중에서 피치를 검출하는 것을 특징으로 할 수 있다.

<33> 앞에서 언급한 기존의 open-loop 피치 검색기의 문제점을 해결하기 위해 본 발명에서 제안하고 있는 피치 검출 방법은 (a) 퍼셉추얼 웨이팅 필터링(PERCEPTUAL WEIGHTING

FILTERING)된 음성신호로부터 정규화된 상관함수를 구하는 단계, (b) 상기 상관도 값을 제공받아서 최대 상관함수 값과 상기 최대 상관함수 값의 후보값들을 검출하는 단계 및 (c) 상기 최대 상관함수와 그 후보값들의 비를 구하고, 소정의 문턱치보다 작은 값을 갖는 최대 상관함수 후보값을 갖는 래그를 피치 후보로 결정하는 단계 및 (d) 상기 음성신호의 이전 프레임의 피치 값을 이용하여 상기 결정된 피치 후보와 최대 상관함수값을 갖는 래그값 중에서 피치를 검출하는 단계를 포함하는 것을 특징으로 한다.

<34> 상기 오픈 루프 피치 검출 방법의 상기 (a) 단계에서 정규화된 상관 함수 $R(d)$ 는 다음 수학식

$$<35> R(d) = \frac{\sum_{n=0}^{N-1} s_w(n-d)s_w(n)}{\sqrt{\sum_{n=2}^{N-1} s_w(n-d)^2}} \quad \text{을 통하여 구하여지고,}$$

<36> 여기서 $R(d)$ 는 상기 정규화된 상관도, $s_w(n)$ 는 perceptual weighting filtering된 상기 음성신호, d 는 래그, d_L 과 d_H 는 최소와 최대 래그, N 은 피치 검색을 위한 윈도우 크기를 의미하며,

<37> 상기 (b) 단계는 상기 정규화된 상관 함수값 중에서 최대 값을 추출하고, 이 최대 값을 찾는 과정에서 나타났던 국부적인 최대값들을 최대값의 후보로 선정하는 것을 특징으로 할 수 있고, 또한 상기 (c) 단계는 상기 최대값의 후보들 각각에 대해 $K(d_x)$ 값을 다음 수학식 $K(d_x) = \text{alpha } K_{\text{lag}}(d_x) + (1 - \text{alpha}) K_{\text{corr}}(d_x)$, $x=1, 2, \dots, 1$ 을 통하여 구한 후 소정의 문턱치보다 작은 $K(d_x)$ 값을 갖는 래그를 피치 후보로 결정하는 것을 특징으로 할 수 있고,

<38> 여기서 alpha 는 소정의 가중치 값,

<39> $K_{lag}(d_x)$ 는 다음 수학식

<40> $K_{lag}(d_x) | [d_{max}/d_x + 0.5] - d_{max}/d_x |$ 을 통하여 구하여지는 값,

<41> d_x 는 상기 후보들 중 어느 하나의 랭크 값,

<42> $K_{corr}(d_x)$ 값은 다음 수학식

<43> $K_{corr}(d_x) = |1 - R(d_{max})/R(d_x)|$ 을 통하여 구하여지는 값,

<44> 1은 d_{max} 이전에 나타났던 최대값 후보들의 수인 것을 특징으로 하며,

<45> 또한 상기 (d) 단계는 이전 프레임의 피치 값을 바탕으로 하여 상기 (c) 단계에서 구한 피치 후보와 최대 상관함수값을 갖는 랭크중에서 피치를 검증 하되 과 가장 가까운 랭크 값을 피치로 검출하는 것을 특징으로 한다.

<46> 이하에서는 본발명을 상세히 설명한다.

<47> 대부분의 음성 코덱에서 사용하는 피치 검색기는 계산의 효율성을 높이기 위해 open-loop 피치 검색기와 closed-loop 피치 검색기로 구성된다. Open-loop 피치 검색기에서는 비교적 간단한 알고리즘으로 대략적인 피치를 구하고, closed-loop 피치 검색기에서는 open-loop 피치 검색기에서 구한 피치 근처의 한정된 랭크값에 대해 합성에의 한 분석 방법을 이용하여 좀 더 정확한 피치를 검색한다. 이렇게 closed-loop 피치 검색기에서는 open-loop 피치 검색에서 구해진 피치값+- alpha의 범위에서 정확한 피치를 검색하기 때문에 만약 open-loop 피치 검색에서 실제 피치의 배수 혹은 역수를 피치로 검출하는 문제가 발생한다면 이는 closed-loop 피치 검색기에서 회복할 수 없는 오류이며 합성된 음성의 품질을 저하시키는 원인이 된다. 본 발명에서 제안하고 있는 open-loop

피치 검색기는 계산량이 적을 뿐만 아니라 배수 혹은 역수 피치 오류의 발생을 최소화 할 수 있는 방법으로 음성 코덱의 음질을 향상시킬 수 있다.

<48> 본 발명이 구현되는 과정을 간략히 설명하면, perceptual weighting filter를 통과 한 음성 신호로부터 미리 정의된 최소와 최대 래그 사이에서 정규화된 상관함수를 계산 한다. 그 후 최대 상관함수 값과 그에 해당하는 래그를 구한다. 그리고 최대 상관함수를 찾는 과정에서 최대 상관함수 후보였던 상관함수 값과 그의 래그를 구한다. 그리고 최대 상관도와 최대 상관도 후보였던 상관도의 비 그리고 래그의 비를 구하고, 정해진 문턱치 보다 작은 값을 갖는 래그를 피치 후보로 결정한다. 최대 상관함수 값을 갖는 래그와 새로이 피치 후보로 결정된 래그 중에서 이전 프레임의 피치와 가까운 래그를 피치로 결정하므로써 본 발명의 목적이 달성된다.

<49> 이하에서는 도면을 참조하여 본 발명의 바람직한 실시예를 상세히 설명한다.

<50> 도 3은 여성 화자가 발성한 음성신호를 perceptual weighting filtering한 신호와 정규화된 상관도를 설명하는 참고도이며, 도 4는 도 3의 d_{max} 와 d_x 값과 그의 상관함수 값이며, 도 5는 본 발명에서 제안하는 Open-loop 피치 검색기의 간략한 블럭도이며, 도 6은 문턱치값을 정하는데 참고하기 위해 구한 최대 상관함수 값을 갖는 래그를 피치로 검출했을 때 배수 피치 오류가 발생한 프레임에 대한 K 값의 분포이고, 도 7은 남성 화자의 음성 신호에 대한 perceptual weighting filtering된 신호와 정규화된 상관도를 설명하는 참고도이며, 도 8은 도 7에 있는 d_x 에 대한 K값을 설명하는 참고도이다. 이하에서 위 도면들은 수시로 참조된다.

<51> 상관함수 계산부(501)는 외부로부터 입력된 perceptual weighting filter를 통과한 음성신호 $s_w(n)$ 을 바탕으로 정규화된 상관함수를 구한다. 상기 정규화된 상관함수 $R(d)$ 는 다음 수학식 1과 같이 구해진다.

<52>

$$R(d) = \frac{\sum_{n=0}^{N-1} s_w(n-d) s_w(n)}{\sqrt{\sum_{n=2}^{N-1} s_w(n-d)^2}}$$

【수학식 1】

<53> 여기서 d 는 래그이며, d_L 과 d_H 는 피치 검색을 위한 최소와 최대 래그이다. 그리고 N 은 피치 검색을 위한 윈도우 사이즈이다. $R(d)$ 는 $s_w(n)$ 과 $s_w(n-d)$ 가 비슷할 때 큰 값을 갖는다. 따라서 만약 $s_w(n)$ 이 주기 P 로 주기적인 신호라면 $R(d)$ 는 P 의 배수마다 피크 값을 갖게 된다. 비록, 래그가 P 일 때 $R(d)$ 가 최대 값을 갖지만, P 의 배수에서 최대값을 갖는 경우도 있다. 이 경우 최대 상관함수값을 갖는 래그를 피치로 선택하면 배수 피치 오류가 발생하게 된다. 특히 남성화자 보다는 피치 주기가 짧은 여성화자의 음성 신호에서 이런 현상이 많이 발생한다.

<54> 도 3은 여성화자의 음성 신호에 대한 perceptual weighting filtering된 과거 신호 $s_w(n-d)$ 와 $R(d)$ 를 나타내고 있다. 피치 검출을 위해서 d_L 부터 d_H 까지 래그를 증가시켜가면서 $R(d)$ 가 최대가 되는 래그 d 를 찾는다. 도 3을 참조하면, d_{max} 에서 $R(d)$ 가 최대가 되지만 만약 d_{max} 를 피치로 검출한다면 실제 피치의 두 배가 되는 래그를 피치로 검출하는 배수 피치 오류가 발생하게 됨을 알 수 있다. 정규화된 상관함수 $R(d)$ 는 피치 주기마다 피크를 갖는데, 이 그림과 같이 실제 피치에서의 상관함수값보다 배수 래그에서의 상관함수가 더 큰 경우에는 배수 피치 오류가 발생한다. 도 3에서 d_1 에서의 상관함수 $R(d_1)$ 은 $R(d_{max})$ 가 최대값으로 선택되기 전까지 최대값이었다.

<55> 도 4에는 래그 d_{-1} 과 d_{\max} 그리고 각각의 상관함수값을 나타내었다. d_{\max} 는 d_{-1} 의 약 2배에 해당하는 래그이고, $R(d_{\max})$ 와 $R(d_{-1})$ 의 차이가 아주 적음을 알 수 있다. 이러한 데이터를 바탕으로 d_{-1} 이 실제 피치일 가능성은 생각할 수 있는데, 본 발명에서는 미리 정의된 최소와 최대 래그에 대해 정규화된 상관함수를 구하고(이는 상관함수 계산부(501)에서 수행된다), 최대의 상관함수를 갖는 래그 d_{\max} 와 $R(d_{\max})$ 가 최대 값으로 검출되기 전까지 최대값으로 나타났던 $R(d_x)$ 와 그의 래그를 검출(이는 최대상관함수 및 래그 검출부(502)에서 수행된다)한 후 피치후보 결정(503)에서는 이들의 비를 계산한 후 소정의 문턱치보다 작은 값을 갖는 최대값 후보이 form를 새로운 피치 후보로 결정한 다음 피치 검출부(504)에서는 이전 프레임의 피치와 새로이 결정된 피치 후보 그리고 최대상관함수값을 갖는 래그를 이용하여 배수 피치 오류를 줄일 수 있는 새로운 open-loop 피치 검출 방법을 제안한다. 이때 d_{\max} 가 실제 피치 혹은 실제 피치의 배수인 경우가 대부분이기 때문에 d_{\max} 가 실제 피치의 배수 래그일 가능성만을 고려한다.

<56> 먼저, 아래와 같이 래그의 비와 상관함수의 비를 이용하여 $K(d_x)$ 를 구한다.

<57> 【수학식 2】 $K(d_x) = \text{alpha} K_{\text{lag}}(d_x) + (1 - \text{alpha}) K_{\text{corr}}(d_x),$

$$x=1, 2, \dots, l$$

<58> 여기서 alpha 는 래그의 비와 상관함수의 비에 주는 가중치 값으로 본 발명에서는 0.5를 사용하였고, l 은 d_{\max} 이전에 나타났던 최대값 후보들의 수이다.

<59> $K_{\text{lag}}(d_x)$ 는 최대 상관함수값을 갖는 래그 d_{\max} 와 d_{\max} 이전에 나타났던 최대값 후보들의 래그 d_x 의 비로 아래와 같이 구한다.

<60> 【수학식 3】 $K_{\text{lag}}(d_x) = [d_{\max}/d_x + 0.5] - d_{\max}/d_x$

<61> 만일 d_{\max} 가 d_x 의 배수 form에 해당하는 값이라면 $K_{\text{lag}}(d_x)$ 는 아주 작은 값을 갖게 된다.

<62> 또한 d_{\max} 와 d_x 의 상관함수 비는 아래와 같이 구할 수 있다.

<63> **【수학식 4】** $K_{\text{corr}}(d_x) = |1 - R(d_{\max})/R(d_x)|$

<64> 앞에서도 언급했듯이 $R(d)$ 는 피치주기의 배수마다 피크를 갖기 때문에 만일 d_{\max}

와 d_x 가 배수 관계를 가지고 있다면 $K_{\text{lag}}(d_x)$ 는 1에 근접한 값이 된다.

따라서 d_{\max} 에서의 상관함수와 d_x 에서의 상관함수값의 차가 작을수록 $K_{\text{corr}}(d_x)$ 은 작아진다. 그러므로 수학식 2의 K값이 작을수록 d_{\max} 가 d_x 의 배수일 확률이 높아진다.

<65> 피치 후보 결정부(503)에서 어떤 문턱치보다 작은 $K(d_x)$ 값을 갖는 d_x 를 피치 랜드 후보로 선택한다. 문턱치는 실험에 의해 구해진 값으로 이 값을 구하기 위해 최대상관함수값을 갖는 랜드를 피치로 검출할 경우에 배수 피치 오류가 발생하는 프레임에 대한 $K(d_x)$ 값의 분포(도 6)를 구하였다. 이 분포를 바탕으로 하여 문턱치값을 0.3으로 선택하였다. 남성화자의 경우 실제 피치의 배수뿐만 아니라 역수에서도 피크가 나타날 수 있다.

<66> 따라서 피치 검출부(504)에서는 실제 피치의 역수 랜드가 피치로 선택되는 역수 피치 오류를 막기 위해 이전 프레임의 피치 값을 이용한다. 따라서 d_{\max} 와 피치 후보 결정부에서 구한 피치 후보 중에서 이전 프레임과의 차이가 적은 랜드를 피치로 정한다.

<67> 도 7은 남성화자의 음성 신호에 대한 perceptual weighting filtering 된 음성신호

$s_w(n-d)$ 와 $R(d)$ 를 보여주고 있다. 이 그림에서 d_1 , d_2 , d_3 는 d_{max} 에서 최대 상관함수값이 검출되기 전까지 최대 값으로 선택되었던 적이 있는 래그 값이다.

<68> 도 8은 이를 래그와 상관함수 그리고 $K(d_x)$ 값을 보여주고 있다. 이 경우 d_{max} 와 $K(d_x)$ 값이 정해진 문턱치 보다 작은 d_3 이 피치 후보가 되는데, 이전 프레임의 피치가 45이므로 d_3 을 피치로 검출한다.

<69> 도 5에서 보여지고 있는 본 발명에 따른 피치 검출장치를 간략히 설명하면 다음과 같다.

<70> 먼저 상관함수 계산부(501)는 perceptual weighting filtering된 음성신호로부터 정규화된 상관함수를 계산한다. 이때 정규화된 정규화된 상관함수 $R(d)$ 는 수학식 1을 통하여 구해진다.

<71> 그 후에 최대상관함수 및 래그 검출부(502)는 위 상관함수 계산부(501)에서 계산된 정규화된 상관함수값을 입력받아서 최대 상관함수값 및 그 래그를 검출하고 위 최대 상관함수의 후보와 그 래그 값을 검출한다.

<72> 피치 후보 결정부(503)는 위에서 검출된 최대 상관함수값 및 그 후보 값의 비, 최대 상관함수의 래그와 그 후보들의 래그의 비 및 소정의 가중치 값을 이용하여 위 후보들 각각에 대응되는 $K(d_x)$ 값을 구하고, 소정의 문턱치(기준값)보다 작은 $K(d_x)$ 값을 갖는 래그를 피치 후보로 결정한다.

<73> 피치 검출부(504)에서는 피치 후보와 최대 상관함수값을 갖는 래그중에서 이전 프레임의 피치와 가장 가까운 래그를 피치로 결정한다.

<74> 한편, 상술한 본 발명의 실시예들은 컴퓨터에서 실행될 수 있는 프로그램으로 작성 가능하고, 컴퓨터로 읽을 수 있는 기록매체를 이용하여 상기 프로그램을 동작시키는 범용 디지털 컴퓨터에서 구현될 수 있다.

<75> 상기 컴퓨터로 읽을 수 있는 기록매체는 마그네틱 저장매체(예를 들면, 롬, 플로피 디스크, 하드디스크 등), 광학적 판독 매체(예를 들면, 씨디롬, 디브이디 등) 및 캐리어 웨이브(예를 들면, 인터넷을 통한 전송)와 같은 저장매체를 포함한다.

<76> 이제까지 본 발명에 대하여 그 바람직한 실시예들을 중심으로 살펴보았다. 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자는 본 발명이 본 발명의 본질적인 특성에서 벗어나지 않는 범위에서 변형된 형태로 구현될 수 있음을 이해할 수 있을 것이다. 그러므로 개시된 실시예들은 한정적인 관점이 아니라 설명적인 관점에서 고려되어야 한다. 본 발명의 범위는 전술한 설명이 아니라 특허청구범위에 나타나 있으며, 그와 동등한 범위 내에 있는 모든 차이점은 본 발명에 포함된 것으로 해석되어야 할 것이다.

【발명의 효과】

<77> 일반적으로 사용하고 있는 CELP 형태의 음성 코덱에서는 한 프레임의 음성신호로부터 음성 신호의 스펙트럼 포락선을 나타내는 LPC 계수와 주기적인 특성을 나타내는 피치 그리고 고정 코드북으로 모델링 되는 예기신호에 대한 정보 등을 추출하고, 다시 이를 정보를 이용하여 음성 신호를 합성한다. 이때 피치 검출 시에 발생할 수 있는 배수 혹은 역수 피치 오류는 합성된 음성의 품질을 저하시키는 중요한 원인이 된다. 정확한 피치 검출은 음성 코덱의 음질 향상에 중요한 역할을 한다. 본 발명에서 제안하고 있는

Open-loop 피치 검색기는 계산량이 적을 뿐만 아니라 기존의 알고리즘에 비해 배수 혹은 역수 피치 오류가 적기 때문에 음성 코덱의 품음 향상에 기여할 수 있다.

【특허청구범위】

【청구항 1】

입력받은 소정의 음성신호의 피치를 검출하는 음성 코덱의 오픈 루프 피치 검출장치에 있어서,

퍼셉추얼 웨이팅 필터링된 음성신호로부터 정규화된 상관함수를 구하는 상관함수 계산부;

상기 상관함수값을 제공받아서 상기 상관함수값들 중 최대 상관함수값과 그에 대응되는 래그 값, 상기 최대 상관함수 값의 후보들과 그들에 대응되는 래그 값들을 검출하는 최대 상관함수 및 래그 검출부;

상기 검출된 최대 상관함수 값과 그 후보들의 값의 비 및 최대 상관함수를 갖는 래그와 그 후보들의 래그들의 비를 이용하여 피치 후보를 결정하는 피치 후보 결정부; 및

및 상기 음성신호의 이전 프레임의 피치를 이용하여 상기 피치 후보와 최대 상관함수를 갖는 래그중에서 피치를 검출하는 피치 검출부를 포함하는 것을 특징으로 하는 오픈 루프 피치 검출 장치.

【청구항 2】

제1항에 있어서, 상기 최대 상관함수 및 래그 검출부는
상기 정규화된 상관함수 값들중에서 최대값을 추출하고, 상기 최대값이 추출되기 전에 최대값으로 선택되었던 값들을 상기 최대값의 후보들로 선정하는 것을 특징으로 하는 오픈 루프 피치 검출 장치.

【청구항 3】

제1항에 있어서, 상기 피치 검출부는

상기 최대값의 후보들 각각에 대응되는 $K(d_x)$ 값을 다음 수학식 $K(d_x) = \alpha K_{\text{lag}}(d_x) + (1 - \alpha) K_{\text{corr}}(d_x)$, $x=1, 2, \dots, l$ 을 통하여 구하고, 소정의 문 턱치보다 작은 $K(d_x)$ 값을 갖는 래그와 최대 상관함수를 갖는 래그중에서 상기 이전 프레임의 피치 값에 가장 가까운 래그 값을 피치로 검출하되,

여기서 α 는 소정의 가중치 값,

$K_{\text{lag}}(d_x)$ 는 다음 수학식

$K_{\text{lag}}(d_x) = |[d_{\text{max}}/d_x + 0.5] - d_{\text{max}}/d_x|$ 을 통하여 구하여 지는 값,

l 은 d_{max} 이전에 나타났던 최대값 후보들의 수,

d_x 는 상기 후보들 중 어느 하나의 래그 값,

$K_{\text{corr}}(d_x)$ 값은 다음 수학식

$K_{\text{corr}}(d_x) = |1 - R(d_{\text{max}})/R(d_x)|$ 을 통하여 구하여 지는 값인 것을 특징으로 하는 오픈 루프 피치 검출 장치.

【청구항 4】

입력받은 소정의 음성신호의 피치를 검출하는 음성 코덱의 오픈 루프 피치 검색장 치에서 피치를 검출하는 방법에 있어서,

(a) 퍼셉추얼 웨이팅 필터링(PERCEPTUAL WEIGHTING FILTERING)된 음성신호로부터 정규화된 상관함수를 구하는 단계;

(b) 상기 상관함수 값들 중 최대 상관함수 값과 그에 대응되는 래그 값, 상기 최대 상관함수 값의 후보들과 그들에 대응되는 래그 값들을 검출하는 단계;

(c) 상기 검출된 최대 상관함수 값과 그 후보들의 값의 비 및 최대 상관함수를 갖는 래그 와 그 후보들의 래그들의 비를 이용하여 피치 후보를 결정하는 단계; 및

(d) 상기 음성신호의 이전 프레임의 피치 값을 입력받아서 피치 후보와 최대 상관함수를 갖는 래그 중에서 피치를 검출하는 단계를 포함하는 것을 특징으로 하는 오픈 루프 피치 검출 방법.

【청구항 5】

제4항에 있어서, 상기 (b) 단계는

상기 정규화된 상관 함수 값중에서 가장 큰 값을 상기 최대값으로, 그리고 최대값이 선택되기 전까지의 래그에서 나타났던 국부적인 최대값들을 최대값의 후보로 선정하는 것을 특징으로 하는 오픈 루프 피치 검출 방법.

【청구항 6】

제5항에 있어서, 상기 (c) 단계는

상기 최대값들의 후보들 각각에 대응되는 $K(d_x)$ 값들을 다음 수학식

$$K(d_x) = \alpha K_{\text{lag}}(d_x) + (1 - \alpha) K_{\text{corr}}(d_x), \quad x=1,2,\dots,1$$
을 통하여 구하고, 상기 후보들 각각에 대응되는 d_{\max} 와 $K(d_x)$ 들 중 소정의 문턱치보다 작은 값을 피치 후보로 결정하되,

여기서 α 는 소정의 가중치 값,

$K_{\text{lag}}(d_x)$ 는 다음 수학식

$K_{lag}(d_x) | [d_{max}/d_x + 0.5] - d_{max}/d_x |$ 을 통하여 구하여 지는 값,

d_x 는 상기 후보들 중 어느 하나의 래그 값,

$K_{corr}(d_x)$ 값은 다음 수학식

$K_{corr}(d_x) = |1 - R(d_{max})/R(d_x)|$ 을 통하여 구하여 지는 값,

I 은 d_{max} 전에 나타났던 최대값 후보들의 수 인 것을 특징으로 하는 오픈 루프 피치 검출 방법.

【청구항 7】

제 5항에 있어서 상기 (d)단계는

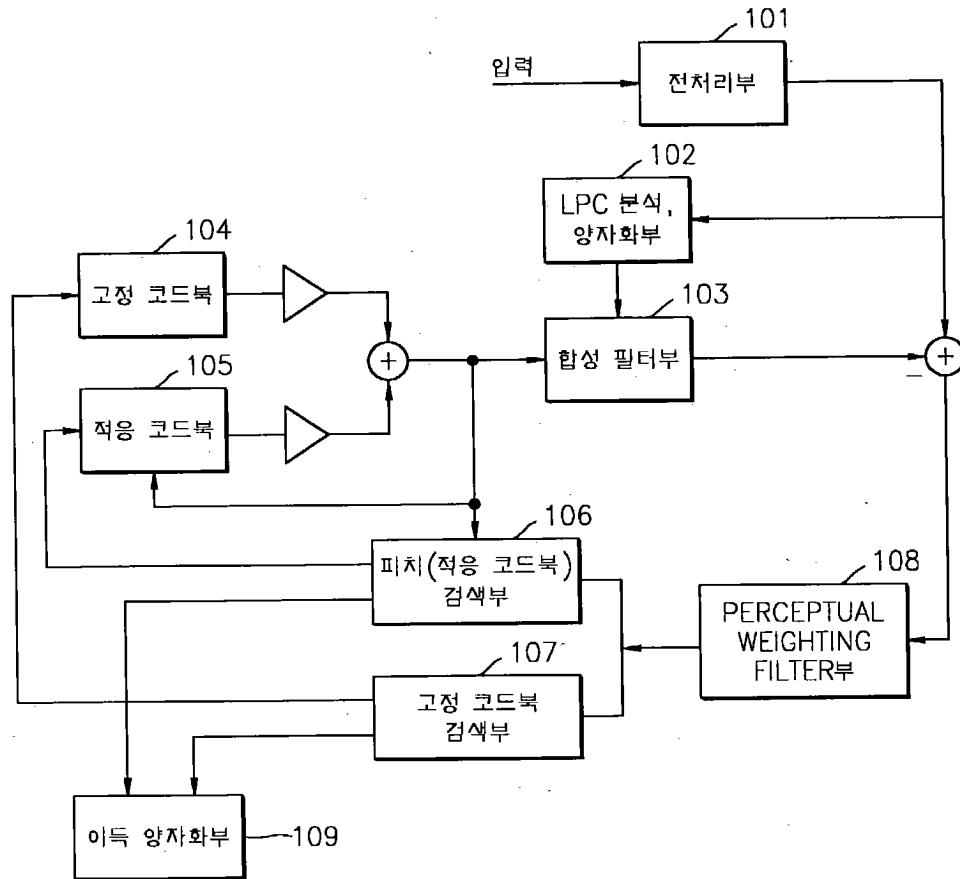
이전 프레임의 피치를 이용하여 여러 피치 후보중에서 이전 프레임의 피치와 가장 가까운 래그값을 피치로 검출하는 오픈 루프 피치 검출 방법.

【청구항 8】

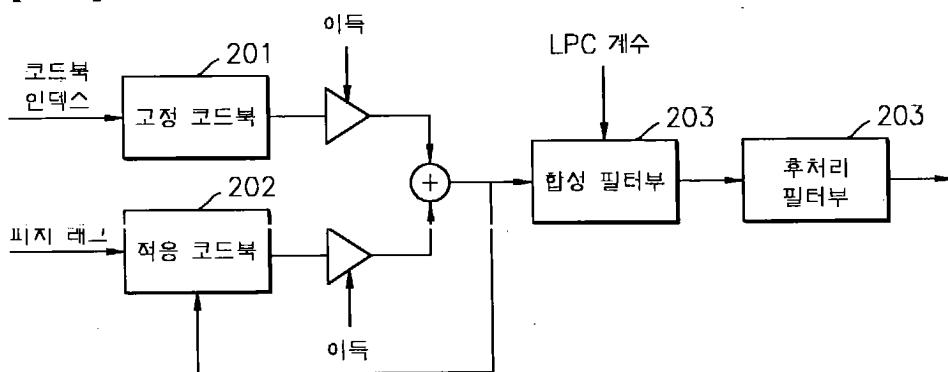
제4항 내지 제8항에 있어서 어느 한 항의 방법을 컴퓨터로 실행시킬 수 있도록 작성된 프로그램을 저장하는 컴퓨터로 읽을 수 있는 기록매체.

【도면】

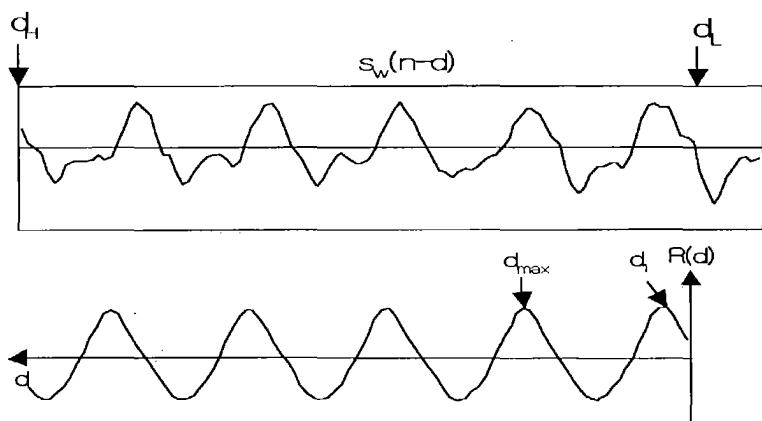
【도 1】



【도 2】



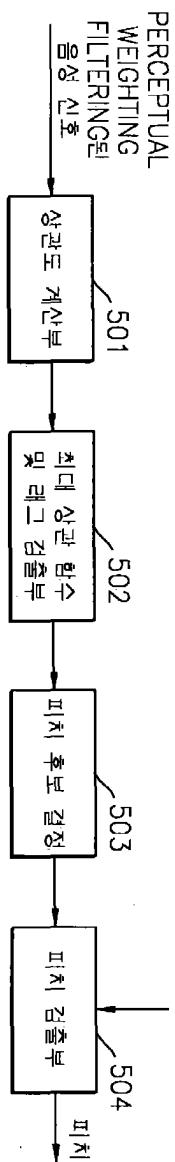
【도 3】



【도 4】

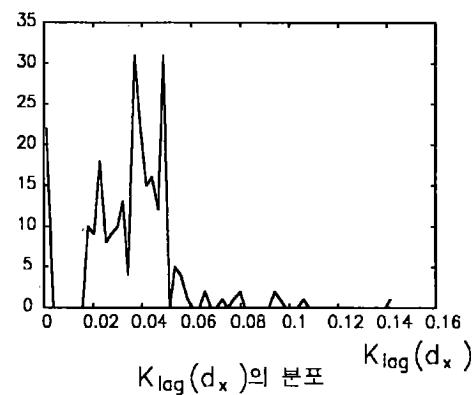
	래그	상관도 $R(d)$
d_{\max}	41	1970
d_1	21	1956

【도 5】

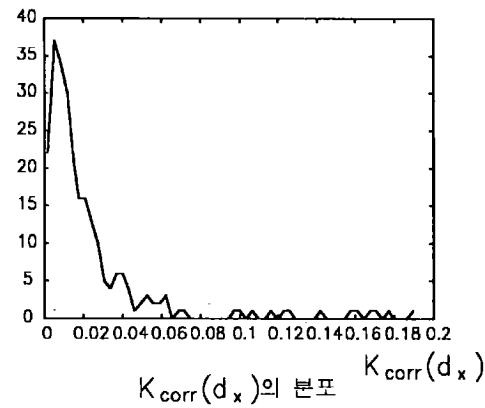


【도 6】

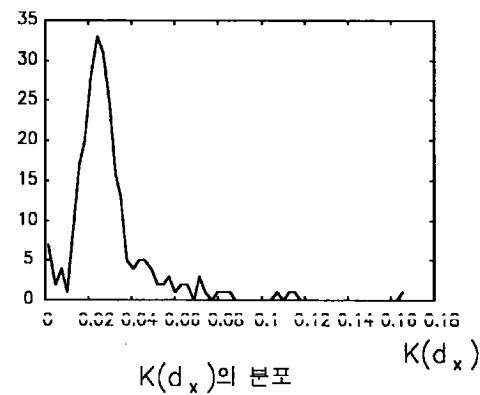
발생 빈도



발생 빈도



발생 빈도

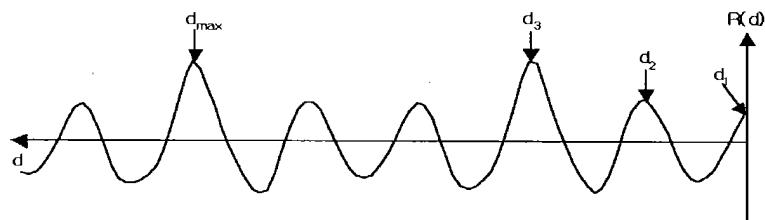
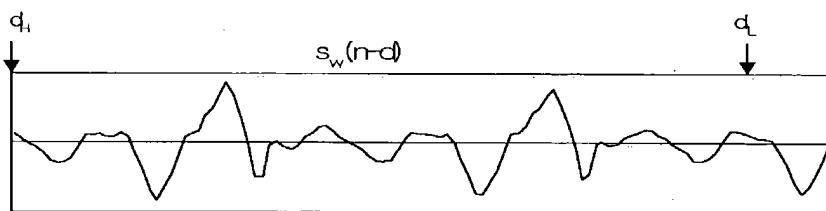




1020020061787

출력 일자: 2003/7/11

【도 7】



【도 8】

	래그	상관도 $R(d)$	$K(d_x)$
d_{\max}	91	6351	
d_s	46	6256	0.0177
d_2	30	3340	0.4674
d_1	17	2133	1.1652